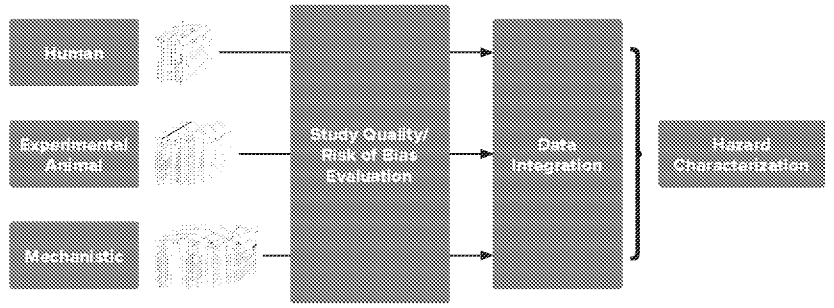
Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

Poster #: P306
Abstract #: 2801

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804. [3]ToxStrategies, Mission Viejo, CA 92691. [4]ToxStrategies, Tallahassee, FL 32309. [5]ToxStrategies, Katy, TX 77494



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characterization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:**[*] 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

[*] OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments
Epidemiology	10	9
Animal Toxicology	11	12
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study] In vivo: 5 In vitro (cell culture): 26 In vitro (in ovo): 21 In vitro (ex ovo): 3 In vitro (ex vivo): 7 In vitro (zebrafish): 5 Unknown model: 1
Total	43	89

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

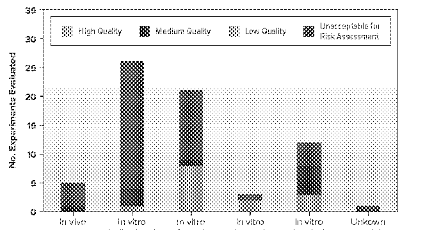


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

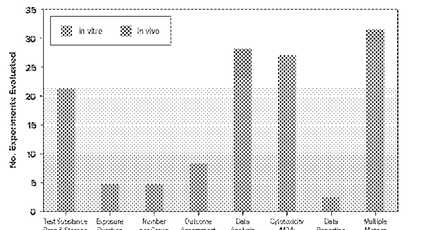
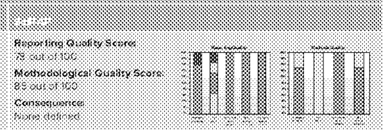
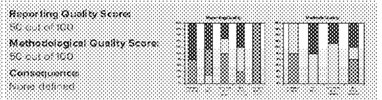
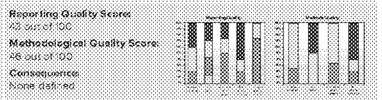


Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Endpoint	Study Quality Evaluation Outcome: TSCA	SciRAP	
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined	
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined	
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined	

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion

Conclusions and Confidence By Stream: Human Data	Animal Data	Mechanistic Data	Integrated Conclusion
Low to very low (inadequate) confidence to determine the potential for, or the direction of, an effect	+ Very high level of confidence supporting lack of association between TCE exposure (oral and inhalation) and CHD Single inconsistency explained by methological and reporting limitations*	+ Low confidence in the body of evidence to characterize CHD or mechanistic events potentially associated with CHD Complex interpretation given inconsistinecy in findings relative to other streams and limitations in relevance of non mammalian study models combined with limitations in study quality	→ Available evidence support a lack of association between gestational TCE exposure and CHD in humans

Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References
Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406647
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529815.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckert PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.

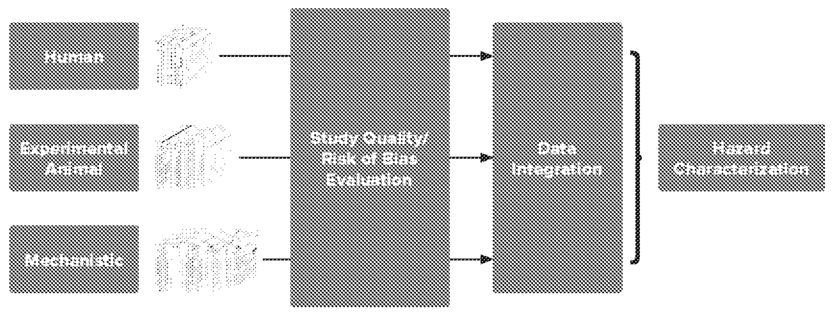
ToxStrategies

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

Poster #: P306
Abstract #: 2801

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804. [3]ToxStrategies, Mission Viejo, CA 92691. [4]ToxStrategies, Tallahassee, FL 32309. [5]ToxStrategies, Katy, TX 77494



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characterization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:**[*] 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

[*] OHAT RoB guidelines encourages reviewers to expend/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments
Epidemiology	10	9
Animal Toxicology	11	12
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study] In vivo: 5; In vitro (cell culture): 26; In vitro (in ovo): 21; In vitro (ex ovo): 3; In vitro (ex vivo): 7; In vitro (zebrafish): 5; Unknown model: 1
Total	43	89

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross-sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

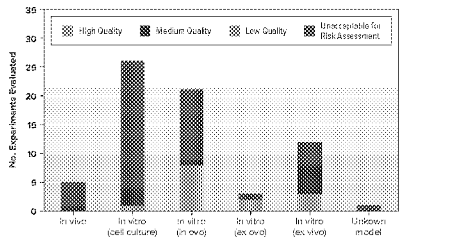


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

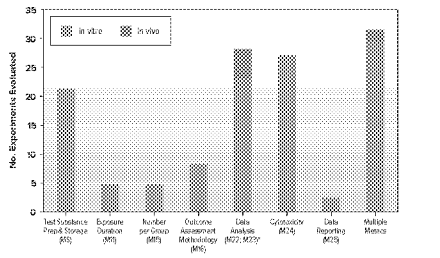
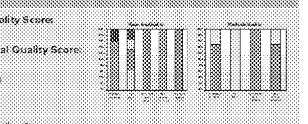
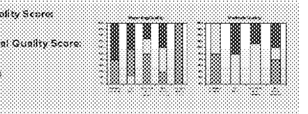
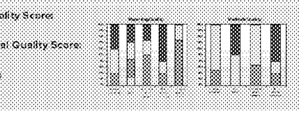


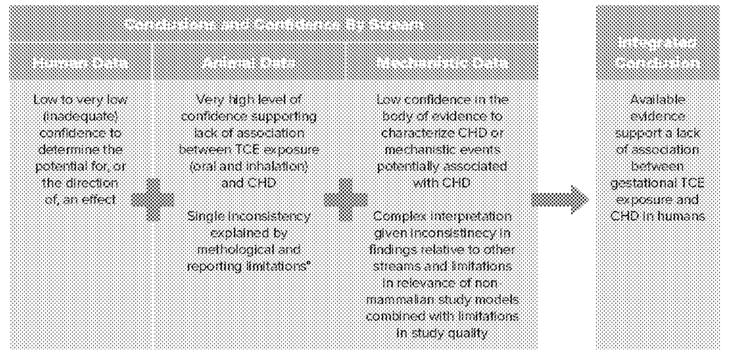
Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Endpoints	Study Quality Evaluation Outcomes: TSCA	SciRAP	
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined	
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined	
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined	

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion



Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References

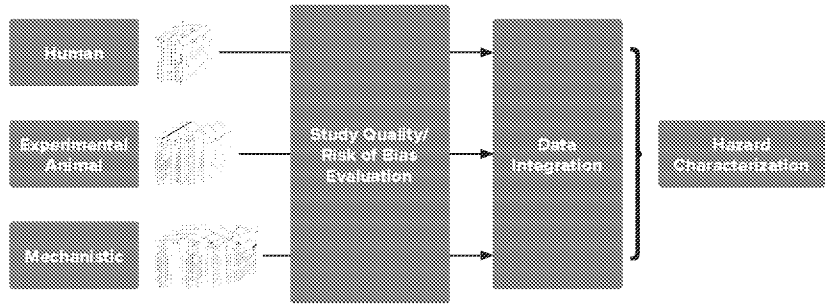
Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406647
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529815.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

Poster #: P305
Abstract #: 2801

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804. [3]ToxStrategies, Mission Viejo, CA 92691. [4]ToxStrategies, Tallahassee, FL 32309. [5]ToxStrategies, Katy, TX 77494



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characterization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:**[*] 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

[*] OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments	
Epidemiology	10	9	
Animal Toxicology	11	12	
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study]	
		In vivo: 5 In vitro (cell culture): 26 In vitro (in ovo): 21 In vitro (ex ovo): 3	In vitro (ex vivo): 7 In vitro (zebrafish): 5 Unknown model: 1
Total	43	89	

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross-sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

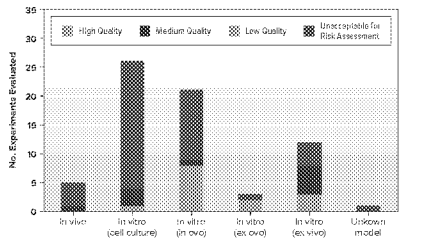


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

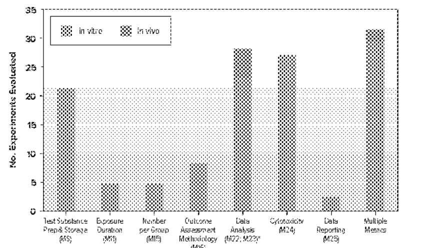


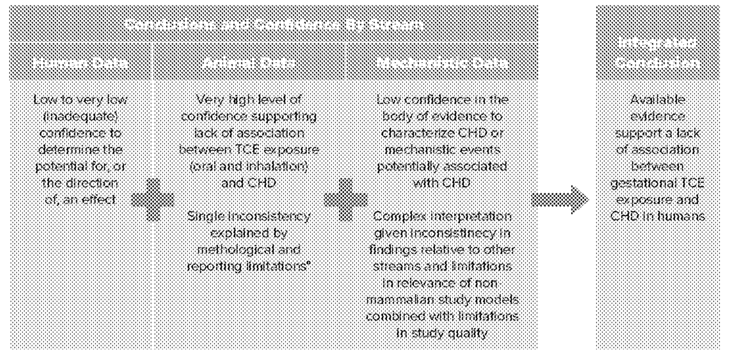
Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Endpoint	Study Quality Evaluation Outcome: TSCA	SciRAP
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion



Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References

Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406647
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529815.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.

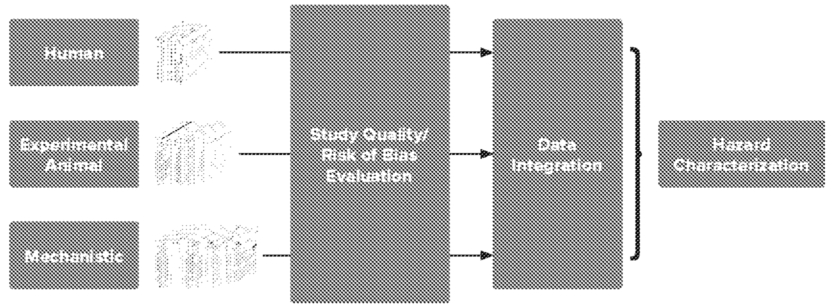
ToxStrategies

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804. [3]ToxStrategies, Mission Viejo, CA 92691. [4]ToxStrategies, Tallahassee, FL 32309. [5]ToxStrategies, Katy, TX 77494

Poster #: P306
Abstract #: 3801



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characterization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:[a]** 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

[a] OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments
Epidemiology	10	9
Animal Toxicology	11	12
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study] In vivo: 5; In vitro (cell culture): 26; In vitro (in ovo): 21; In vitro (ex ovo): 3; In vitro (ex vivo): 7; In vitro (zebrafish): 5; Unknown model: 1
Total	43	89

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

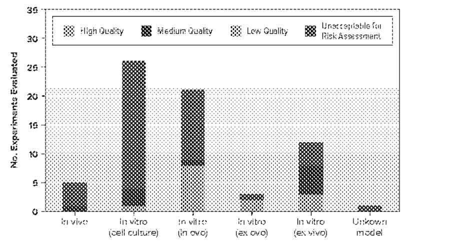


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

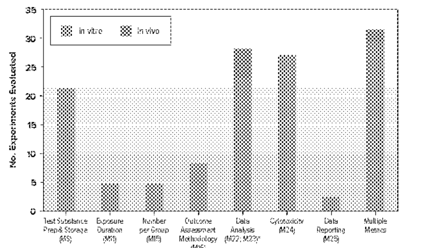


Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Endpoint	Study Quality Evaluation Outcome: TSCA	SciRAP
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion

Conclusions and Confidence By Stream: Human Data	Animal Data	Mechanistic Data	Integrated Conclusion
Low to very low (inadequate) confidence to determine the potential for, or the direction of, an effect	Very high level of confidence supporting lack of association between TCE exposure (oral and inhalation) and CHD Single inconsistency explained by methological and reporting limitations*	Low confidence in the body of evidence to characterize CHD or mechanistic events potentially associated with CHD Complex interpretation given inconsistinecy in findings relative to other streams and limitations in relevance of non mammalian study models combined with limitations in study quality	Available evidence support a lack of association between gestational TCE exposure and CHD in humans

Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References
Bove FJ. (1996). PMID: 8794638.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406847
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288582
Narotsky MG et al. (1995). PMID: 8529816.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.

ED_004056A_00217279-00001

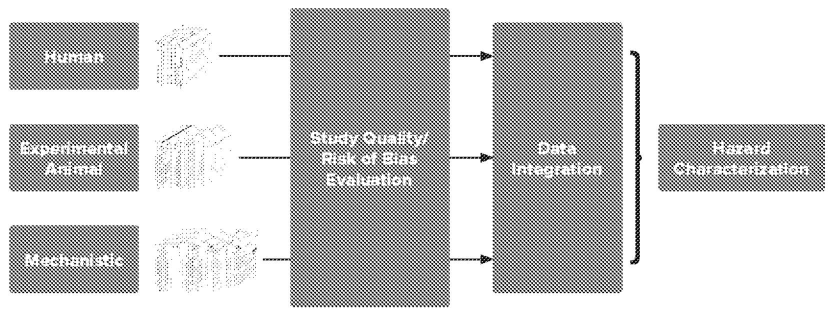
ToxStrategies

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

[1]ToxStrategies, Austin, TX 78759; [2]ToxStrategies, Asheville, NC 28804; [3]ToxStrategies, Mission Viejo, CA 92691; [4]ToxStrategies, Tallahassee, FL 32309; [5]ToxStrategies, Katy, TX 77494

Poster #: P306
Abstract #: 2801



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2016)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characerization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:*** 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

* OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments	
Epidemiology	10	9	
Animal Toxicology	11	12	
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study]	
		In vivo: 5 In vitro (cell culture): 26 In vitro (in ovo): 21 In vitro (ex ovo): 3	In vitro (ex vivo): 7 In vitro (zebrafish): 5 Unknown model: 1
Total	**43**	**89**	

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation†	OHAT RoB Designation*
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross-sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

† For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
* OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation†	OHAT RoB Designation*
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

† For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
* OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

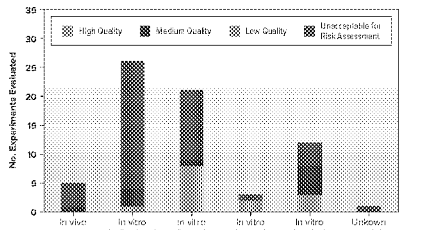


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

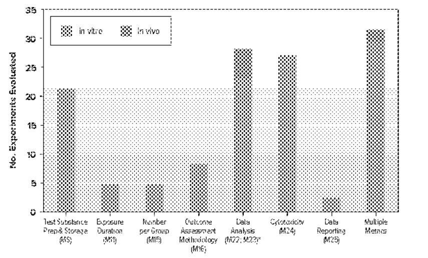


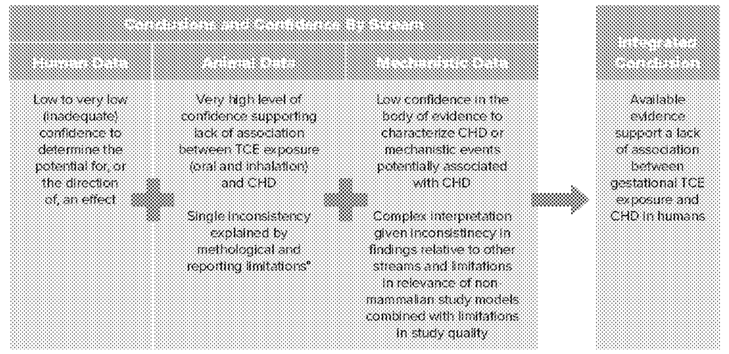
Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	[illegible]	Study Quality Evaluation Outcome: TSCA	SciRAP
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated; multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion



Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References

Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406647
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529815.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2016). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29367719.

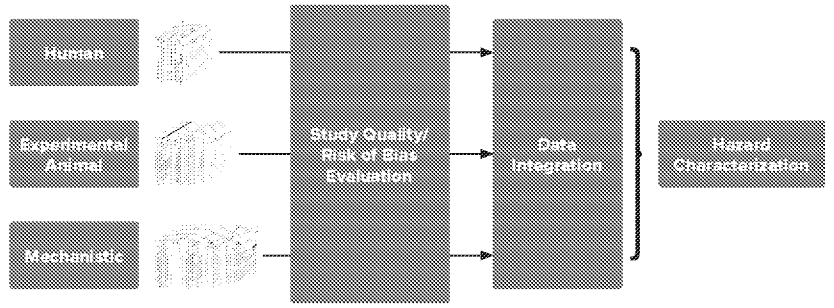
ToxStrategies

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

[1]ToxStrategies, Austin, TX 78759. [2]ToxStrategies, Asheville, NC 28804. [3]ToxStrategies, Mission Viejo, CA 92691. [4]ToxStrategies, Tallahassee, FL 32309. [5]ToxStrategies, Katy, TX 77494

Poster #: P306
Abstract #: 2801



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo, ex ovo, ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characterization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:*** 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

* OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments	
Epidemiology	10	9	
Animal Toxicology	11	12	
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study]	
		In vivo: 5 In vitro (cell culture): 26 In vitro (in ovo): 21 In vitro (ex ovo): 3	In vitro (ex vivo): 7 In vitro (zebrafish): 5 Unknown model: 1
Total	43	89	

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

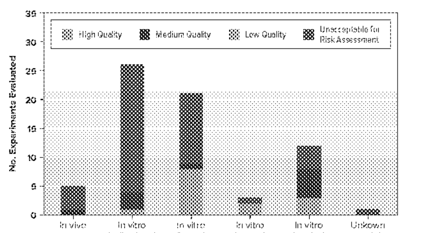


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

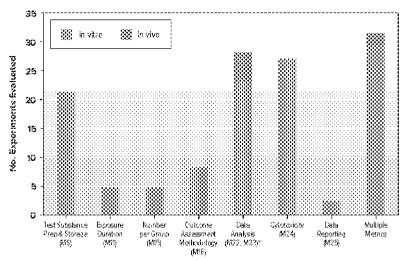


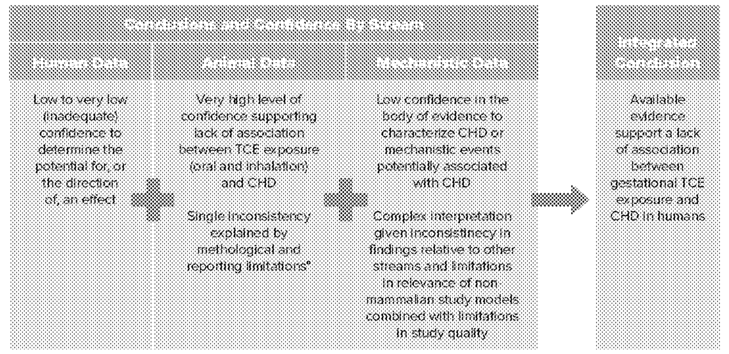
Table 6. Comparison of In Vitro Study Quality Evaluation Tools

[illegible]	[illegible]	Study Quality Evaluation Outcomes: TSCA	SciRAP
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 83 out of 100 **Consequence:** None defined
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion

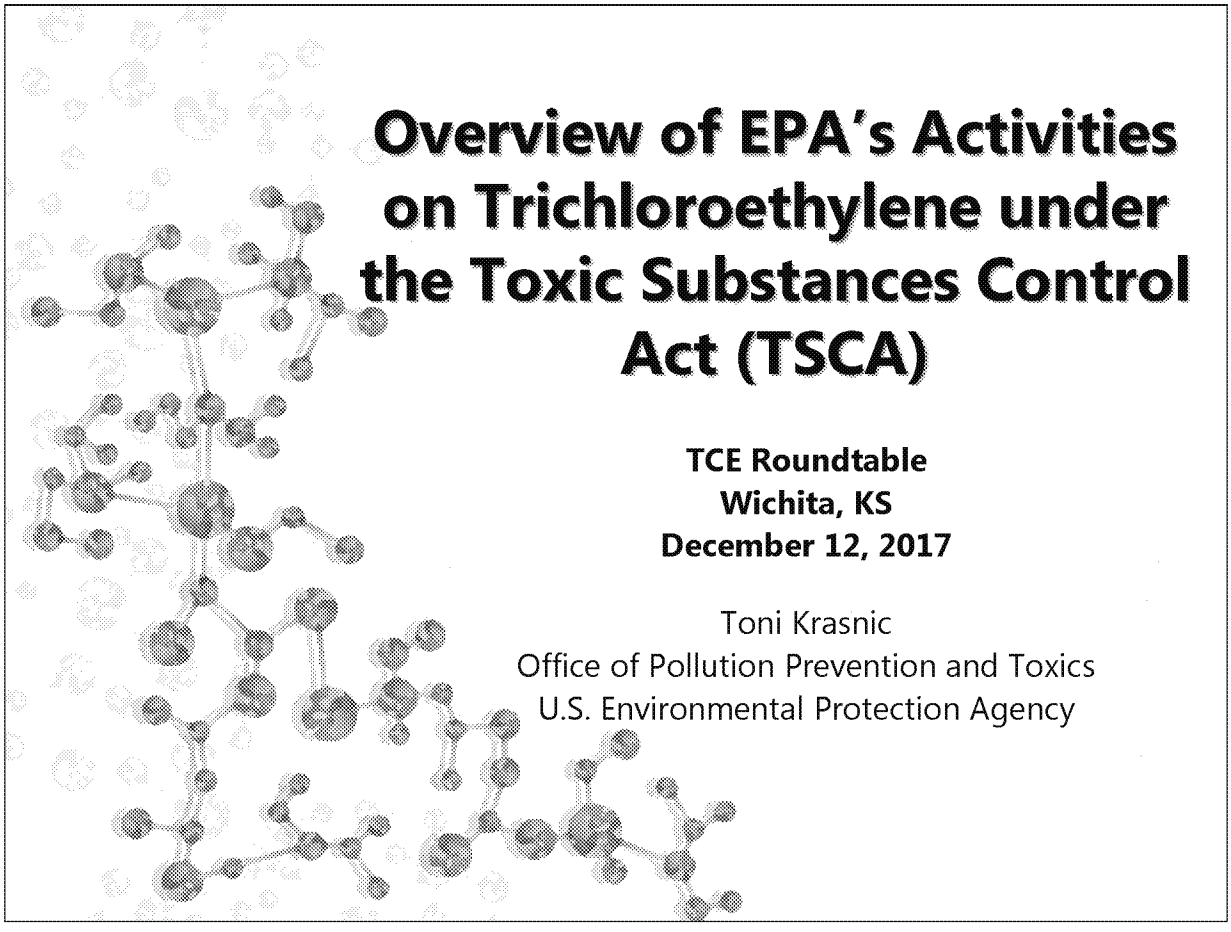


Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References
Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406847
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529816.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.



Overview of EPA's Activities on Trichloroethylene under the Toxic Substances Control Act (TSCA)

TCE Roundtable
Wichita, KS
December 12, 2017

Toni Krasnic
Office of Pollution Prevention and Toxics
U.S. Environmental Protection Agency

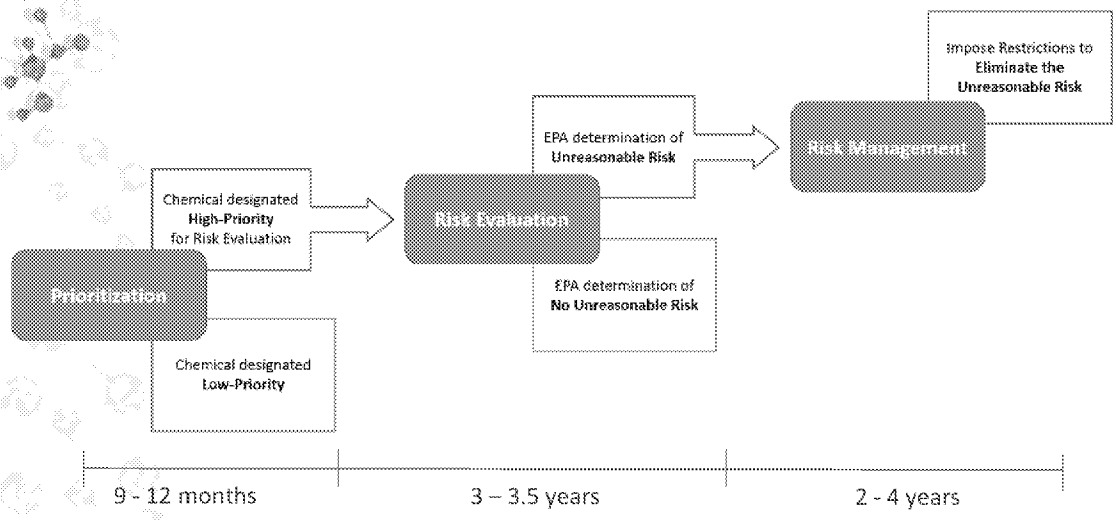
Overview

- The Frank R. Lautenberg Chemical Safety for the 21st Century Act
- Evaluating Risks of Existing Chemicals
- Initial 10 Chemicals
- TCE Overview
- TSCA §6 Rulemakings
- Important Dates
- Additional Information

The New Chemical Safety Law

- The "Frank R. Lautenberg Chemical Safety for the 21st Century Act" was signed by the President going into immediate effect on June 22, 2016
- Amends and updates the Toxic Substances Control Act of 1976
- Passed by large bipartisan margins in the U.S. House (403 to 12) and unanimously in Senate
- Received support from chemical industry and downstream users of chemicals, NGOs and other stakeholders

Evaluating Risks of Existing Chemicals



Risk Evaluation:

Statutory Requirements

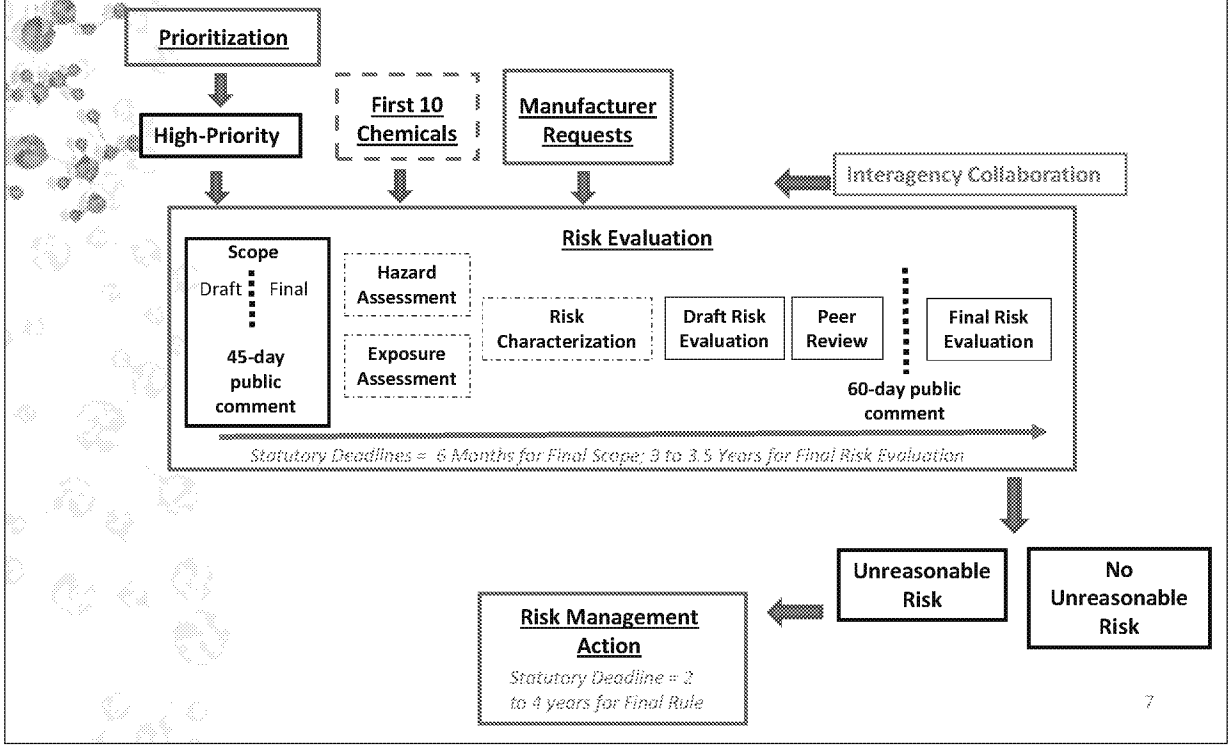
- EPA must establish by rule a process for risk evaluation
  - Risk evaluations will determine if a chemical presents an unreasonable risk of injury to health or the environment under the conditions of use
  - Without consideration of cost or other non-risk factors
  - Including unreasonable risk to potentially exposed or susceptible subpopulation(s)
- This process must be completed within 3 years with a potential 6 month extension
- For each risk evaluation completed, EPA must designate a new high-priority chemical
- Within 3 years, EPA must have initiated 20 high priority chemicals for risk evaluation
  - Additional risk evaluations may come from manufacturer requests

Risk Evaluation:

Statutory Requirements

- Chemicals evaluated under this process
  - First 10 Chemicals from the 2014 Update to Work Plan Chemicals (announced by Dec 19, 2016)
  - High-Priority Designated Chemicals
- Scope documents – published within 6 months of initiation
  - Must identify hazards, exposure, conditions of use, potentially exposed or susceptible subpopulation(s)
- Draft Risk Evaluation
  - Integrate and assess available information on hazards and exposures for the conditions of use of the chemical, including information on specific risks of injury to health or the environment and information on potentially exposed or susceptible subpopulations
  - Publication in Federal Register
  - 30-day public comment
- Final Risk Evaluation – published within 3 years of initiation
  - Publication in Federal Register

Risk Evaluation Process and Timeline



Initial 10 Chemicals

- EPA identified the initial 10 chemicals and formally initiated risk evaluations on December 19, 2016
  - Statute required chemical be drawn from the 2014 Update to TSCA Work Plan
  - Methodology involved screening for hazard, exposure, persistence, and bioaccumulation

1,4-Dioxane
1-Bromopropane
Asbestos
Carbon Tetrachloride
Cyclic Aliphatic Bromide Cluster (HBCD)
Methylene Chloride
N-methylpyrrolidone (NMP)
Pigment Violet 29
Tetrachloroethylene, or perchloroethylene (perc)
Trichloroethylene (TCE)

TCE: Overview

- Volatile organic compound (VOC) and hazardous air pollutant (HAP) classified as a human carcinogen
- Widely used in industrial and commercial processes; there are some uses in consumer products
- More than 255 million lbs per year used in the United States
  - Majority of TCE (~84%) used as an intermediate for manufacturing refrigerant chemicals
  - Much of the remainder used as a solvent for metal degreasing (~15%).
  - A small percentage (~1%) used in other applications, including dry cleaning and consumer uses
  - EPA prioritized assessment of degreasing and other uses, because refrigerant uses take place in enclosed systems where exposures are expected to be comparatively low

TCE: Uses

- Industrial and commercial uses include:
  - Degreaser and cleaner (vapor, cold cleaning, aerosol)
  - Intermediate in refrigerant manufacture
  - Adhesives
  - Sealants
  - Lubricants
  - Die fluid
  - Mold release
  - Spot cleaning in dry cleaning facilities.
- Commercial and consumer uses include:
  - Cleaning wipes
  - Carpet cleaners
  - Adhesives
  - Hoof polish
  - Pepper spray
  - Sealants
  - Lubricants
  - Toner aids

TCE: Risk Assessment

- Final IRIS Health Assessment: 2011
  - Carcinogenic to humans with mutagenic mode of action.
  - Evidence for multiple non-cancer end-points:
    - Kidney, liver, immune system, central nervous system, reproductive, and developmental toxicity.
  - Fetal cardiac malformations specifically identified as a developmental hazard. Hazard conclusion supported by two expert review panels (NRC/NAS-2006, SAB, 2011).
  - See http://cfpub.epa.gov/ncea/iris2/chemicalLanding.cfm?substance_nmbr=199.

TCE: Risk Assessment

- Final TSCA Work Plan Chemical Risk Assessment: July 2014
  - Followed Agency peer review process of publishing a public draft, peer review, and response to peer review and public comment
  - Cancer and non-cancer risks from long-term (chronic) exposure (workers):
    - Most of the occupational exposure scenarios exceeded the target cancer risk range (10^{-6} to 10^{-4}).
    - Non-cancer risks to workers were determined for a range of human health effects.
  - Non-cancer risks identified from short-term (acute) exposure:
    - TCE can irritate the respiratory system and skin and induce central nervous system effects such as light-headedness, drowsiness, and headaches.
    - Developmental effects (i.e., cardiac defects to fetal death).
  - See http://www.epa.gov/assessing-and-managing-chemicals-under-tsca/assessments-tsca-work-plan-chemicals#tce

TSCA §6(a)

- Provides EPA with the authority to prohibit or limit the manufacture, processing, distribution in commerce, use or disposal of a chemical or mixture

- EPA must:
  - Determine after risk evaluation whether a chemical substance or mixture "presents an unreasonable risk of injury to health or the environment, without consideration of costs or other non-risk factors, including an unreasonable risk to a potentially exposed or susceptible subpopulation⋯.under the conditions of use"
  - Apply one or more of the regulatory options to the extent necessary so that the chemical substance no longer presents such risk

Options Under TSCA §6(a)

- Prohibit or otherwise restrict manufacture, processing or distribution in commerce
- Prohibit or otherwise restrict for particular use or above a set concentration
- Require minimum warnings and instructions.
- Require recordkeeping or testing
- Prohibit or regulate manner or method of commercial use
- Prohibit or regulate manner or method of disposal
- Direct manufacturers/processors to give notice of risk to distributers and users and replace or repurchase

TCE: TSCA §6 Rulemakings

- TSCA gave EPA specific authority under TSCA section 26(l)(4) to proceed with risk management actions on chemicals already assessed under the TSCA Work Plan, even though these assessments were for subsets of the chemicals' uses
  - "…the Administrator may publish proposed and final rules under section 6(a) that are consistent with the scope of the completed risk assessment for the chemical substance and consistent with other applicable requirements of section 6."
- Assessments for TCE demonstrated significant risks
  - TCE 1 (spot cleaning and aerosol degreasing)
    - Proposed rule December 16, 2016
    - Comment period closed on March 16, 2017
  - TCE 2 (vapor degreasing)
    - Proposed rule January 19, 2017
    - Comment period closed on May 19, 2017

TCE: Overview of the Proposed Regulations

- EPA has proposed regulations under §6(a) of TSCA to reduce or eliminate risks posed by trichloroethylene (TCE) for the following uses:
  - Use of TCE in aerosol degreasing
    - Consumer and commercial use of TCE in aerosol degreasing
    - Examples:
      - Repair shops: shops that repair automobiles, motorcycles, bicycles, or electronics
      - Fabrication shops: shops that fabricate metals, glass, or plastic components onto a final product (e.g., facilities that produce automobiles, aircraft, appliances, motor vehicle parts, machine parts, and jewelry)
      - Metal plating shops: shops that plate, or coat, metal onto a surface
      - Electronics assembly shops: shops that assemble a wide variety of electronic devices onto circuit boards
  - Use of TCE for spot cleaning in dry cleaning facilities
    - Pre-spotting to remove stains or spots before cleaning the garment in the machine
  - Use of TCE in vapor degreasing
    - All types of vapor degreasing
    - Examples:
      - Vapor degreasing of small parts; fabrication of metal products; instruments and related products; machinery; electrical and electronic equipment; Miscellaneous manufacturing industries

Developing Potential Regulatory Options

- Many options analyzed, including:
  - *Material substitution*: Reducing the concentration of TCE in the degreasing formulation, with concentrations varying from 5 to 95 weight percent.
  - *Equipment substitution* (ES): Replacing open-top vapor degreasing units with an enclosed system to reduce the escape of TCE vapors into the air, which EPA assumes achieves a 98 percent reduction effectiveness.
  - *Engineering controls*: Using local exhaust ventilation (LEV) which is assumed to achieve 90 percent reduction effectiveness and to improve ventilation near worker activity.
  - *Personal protective equipment (PPE)*: Workers and occupational bystanders wearing respirators with an assigned protection factor (APF) varying from 10 to 10,000.
- Combinations of options were also analyzed

Proposed Regulatory Options: Aerosol Degreasing and Spot Cleaning in Dry Cleaning Facilities

- Use of TCE for aerosol degreasing:
  - Prohibit manufacturing, processing, and distribution of TCE for this use; prohibit this commercial use; and require downstream notification when distributing TCE for other uses
- Use of TCE for spot cleaning in dry cleaning facilities:
  - Prohibit manufacturing, processing, and distribution of TCE for this use; prohibit this commercial use; and require downstream notification when distributing TCE for other uses

Proposed Regulatory Option and Request for Comment on Alternative Option: Vapor Degreasing

Proposed regulatory option:

- Prohibit the manufacturing, processing, distribution in commerce of TCE for use in vapor degreasing, prohibit the use of TCE in vapor degreasing, and require downstream notification and recordkeeping.

Alternative option and request for comment:

- Allow use of TCE in certain closed vapor degreasing systems with appropriate personal protective equipment (supplied air respirator of APF 10,000) with the option of meeting an air exposure level.
  - APF is the workplace level of respiratory protection that a respirator or class of respirators is expected to provide to employees. For example, APF 10,000 reduces the exposure concentration by 10,000 times.
  - Risks are reduced so that MOEs are above target benchmarks and cancer is above a risk level of 10^{-6}

Rationale for Proposed Regulatory Option

- Option necessary to ensure the chemical no longer presents an unreasonable risk
  - Eliminates risks for workers and occupational bystanders
  - Regulates the stream of commerce for the specific uses, resulting in increased effectiveness of prohibition
  - Provides awareness of prohibition through downstream notification thus streamlining compliance
  - Switching to alternative methods or chemicals is less costly than adopting and implementing a PPE program
- Other options either did not protect against the unreasonable risk or as is the case with the alternative option were considered more burdensome

Important Dates

- Scope documents were published on June 22, 2017
  - Includes the hazards, exposures, conditions of use, and potentially exposed or susceptible subpopulations that the Administrator expects to consider.
- EPA will publish Problem Formulation documents by Jan. 2018
  - These will further refine the scope, specifically the conditions of use considered in the risk evaluation
- Draft Risk Evaluations
  - Risk evaluations will determine if a chemical presents an unreasonable risk of injury to health or the environment under the conditions of use.
  - Each draft risk evaluation will be peer reviewed.
  - 60-day public comment period on the draft risk evaluation
- EPA will publish Final Risk Evaluations by December 2019 (with a potential 6 month extension)

Additional Information

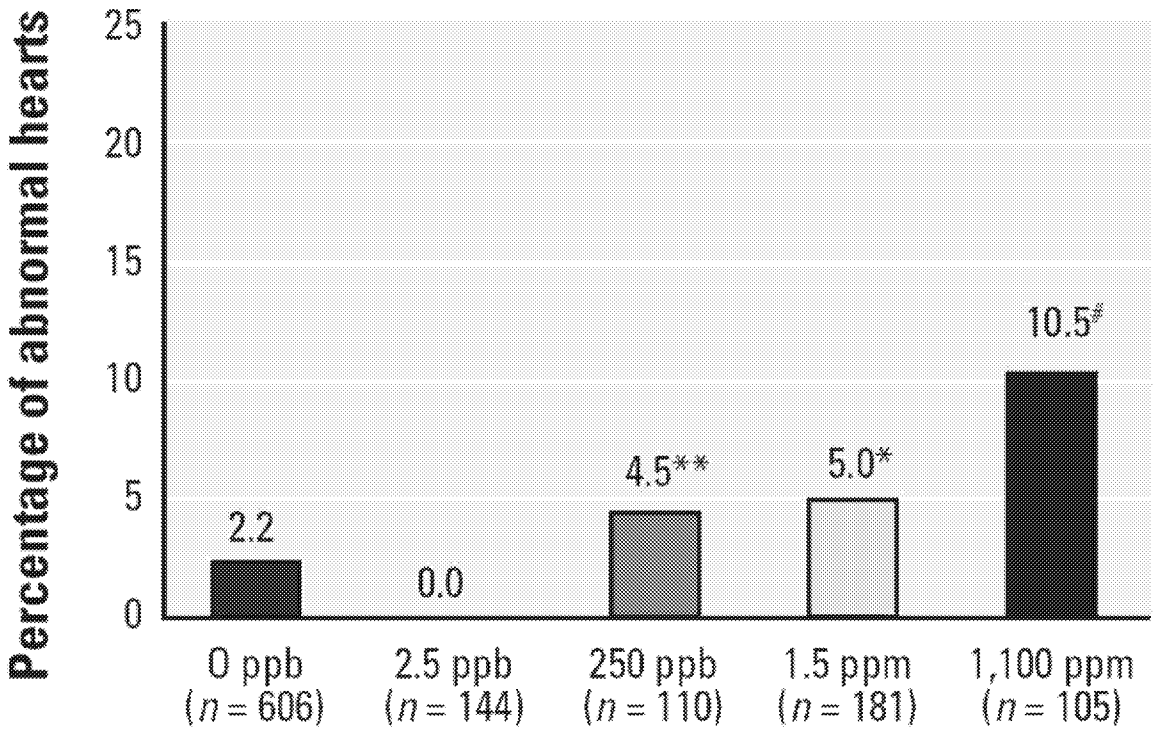
- Contact:
  - Toni Krasnic, 202-564-0984, krasnic.toni@epa.gov
- EPA Website
  - www.epa.gov/assessing-and-managing-chemicals-under-tsca/trichloroethylene-tce
  - EPA's TSCA Implementation Activities website at https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/frank-r-lautenberg-chemical-safety-21st-century-act
- Dockets (www.regulations.gov)
  - Trichloroethylene; TSCA Review and Scoping: EPA-HQ-OPPT-2016-0737
  - TCE Rulemaking under SCA Section 6(a) for Use of TCE in Vapor Degreasing: EPA-HQ-OPPT-2016-0387
  - Rulemaking under TSCA Section 6(a) for Use of TCE in Aerosol Degreasing and Spot Cleaning in Dry Cleaning Facilities: EPA-HQ-OPPT-2016-0163
  - Significant New Use Rule (SNUR) for TCE in Consumer Products: EPA-HQ-OPPT-2014-0697
  - TCE Workshop (July 29-20, 2014): EPA-HQ-OPPT-2014-0327
  - TCE Risk Assessment: EPA-HQ-OPPT-2012-0723

Fetal Cardiac Findings in Rats Exposed to Trichloroethylene (TCE) in Drinking Water

James S. Bus, Exponent, Inc.
On behalf of Halogenated Industry Solvents Alliance (HSIA)

ARA Workshop
Beyond Science and Decisions: From Problem Formulation to Risk Assessment
February 26 & 27, 2019
Texas Commission on Environmental Quality
Austin, Texas

Issue: Increased Cardiac Malformations in Rats Treated with TCE in Drinking Water (Johnson *et al.*, 2003)



From: Johnson *et al., Env. Health Perspect.* 111: 289-292 (2003)

Concerns with Johnson *et al.* (2003)

- 1.5 and 1,100 ppm data published earlier (Dawson *et al.,* 1993).
- Highly unusual dose-response: positive responses over 4,400X dose range.
- Subject to two *errata* (Johnson *et al.,* 2005, 2014) and one letter-to-the-editor correction (Johnson *et al.*, 2004).
- Small study size: 9-13 treated dams per dose.
- Likely lack of concurrent controls preventing matching of per litter incidence treated responses with concurrent control incidence data.
- Non-standard cardiac evaluation: fixation/dissection with manipulation of heart to assess valvular function; technique changed with time.
- Raw data not available for regulatory or public review.

Implications of Use of Johnson *et al.* (2003) for Regulatory Evaluation, e.g., RfD, RfC development

Regulatory Implications – EPA IRIS (2014)

- Based on Johnson study, EPA TCE IRIS set RfC = 0.4 ppb and RfD = 0.0005 mg/kg/day.
- Indoor air exposure exceedances are primarily due to vapor intrusion.

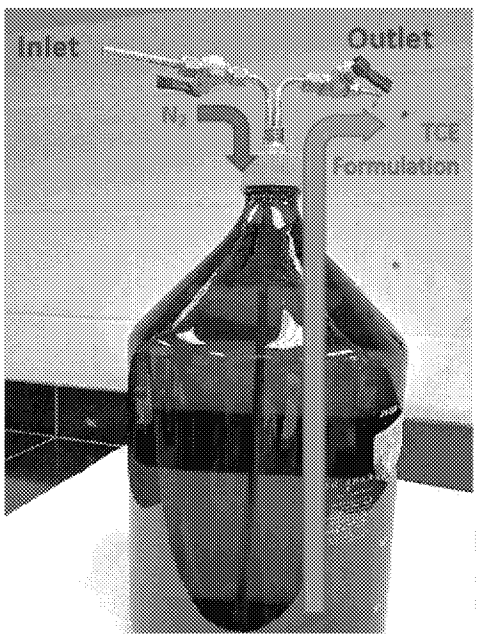
Reproducibility Implications – Other GLP-Quality Studies

- Negative findings with oral gavage TCE (500 mg/kg/day) or TCE oxidative metabolite, TCA (300 mg/kg/day (Fisher *et al.*, 2001)
  - Included Johnson as co-investigator, using Johnson exam techniques
- Negative findings with inhalation TCE (500 ppm) (Carney *et al.*, 2006)

HSIA Drinking Water Repeat Study

- Drinking water doses similar to Johnson *et al.* (2003): 250, 1,500, 500 and 1,000 ppm TCE (1,000 ppm just below TCE water solubility limit)
- 24 pregnant Sprague-Dawley rats per dose group, exposed to TCE in drinking on Gestation Days 1-21.
- Detailed attention to TCE drinking water concentrations due to volatility concerns.
- Focused attention on cardiac evaluations (fresh dissection).
- Retinoic acid used as positive control.
- TCE and TCA determined at various times in maternal and fetal blood.

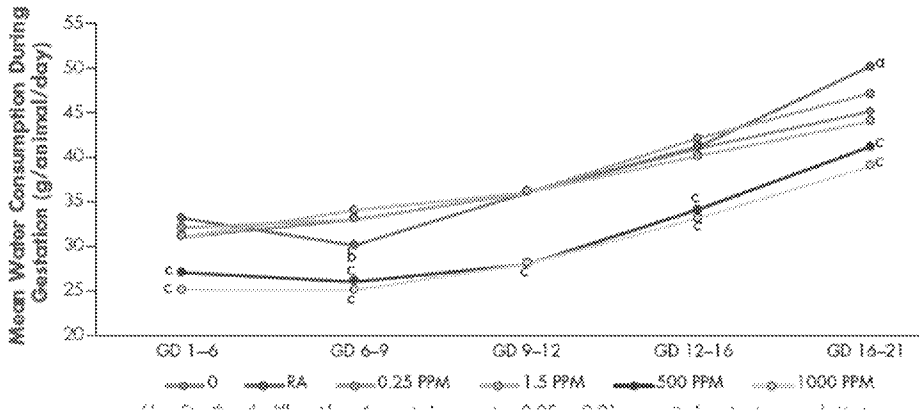
Drinking Water Dose Preparation and Confirmation

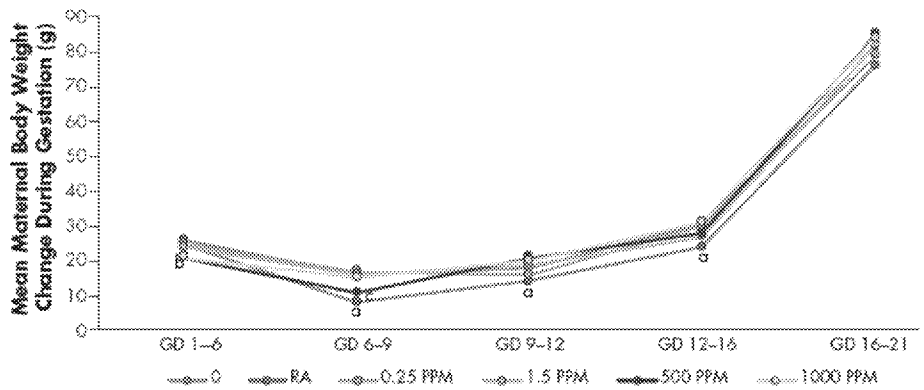


TCE concentrations (% target) in drinking water at:

- Dose preparation: 90 – 130%
- Cage bottle (initial): 94 -- 166%
- Cage bottle (24 hr): 32 – 49%

TCE Decreased Dam Drinking Water Consumption but Body Weights are Similar: Non-Adverse Finding





TCE Drinking Water Treatment did not Affect Ovarian and Uterine Parameters

TCE (ppm)	No. Pregnant Females	Mean No. Corpora Lutea	Mean No. Implantation sites	Pre-Implantation Loss (%)	Resorptions		Dead Feluses	Post Implantation Loss (%)	Mean No. Viable Feluses	Mean Fetal Weights (g)
					Early	Late				
0	24	15.0 ± 2.7	13.7 ± 4.4	1.3 ± 2.2	0.8 ± 1.2	0.0 ± 0.2	0	0.9 ± 1.4	12.8 ± 4.3	6.0 ± 0.3
0.25	23	15.6 ± 2.6	13.2 ± 3.4	2.4 ± 4.7	1.2 ± 2.5	0.0 ± 0.0	0	1.2 ± 2.5	12.0 ± 4.1	6.2 ± 0.3
1.5	24	15.1 ± 2.1	13.8 ± 3.2	1.3 ± 1.8	0.5 ± 0.9	0.0 ± 0.0	0	0.5 ± 0.9	13.4 ± 3.1	5.9 ± 0.4
500	24	15.4 ± 1.9	14.4 ± 1.8	1.0 ± 1.2	0.7 ± 1.1	0.0 ± 0.0	0	0.7 ± 1.1	13.8 ± 2.2	6.0 ± 0.3
1,000	24	16.3 ± 2.2	14.9 ± 2.5	1.3 ± 2.2	0.6 ± 0.7	0.0 ± 0.2	0	0.7 ± 0.8	14.3 ± 2.4	5.9 ± 0.3

Data are presented as mean ± standard deviation, where appropriate.

TCE in Drinking Water did not Increase Cardiac Ventricular Septal Defects (VSDs)

Fetal Parameter	TCE Concentration					Positive Control
	0 ppm	0.25 ppm	1.5 ppm	500 ppm	1000 ppm	RA 15 mg/kg/day
Available Fetuses (Litters)	308 (24)	275 (22)*	321 (24)	330 (24)	342 (24)	269 (25)
Affected Fetuses (Litters)	7 (5)	4 (4)	5 (3)	13 (8)	12 (6)	112 (23)
Mean Litter Proportion (% per Litter)	2.4	1.4	1.5	3.8	3.7	42.2**
No. Fetuses (Size of Opening)	6 (<1mm) 1 (1–2mm)	All (<1mm)	All (<1mm)	All (<1mm)	All (<1mm)	103 (<1 mm) 8 (1–2 mm) 1 (>2 mm)
Location of Opening	Mem	Mem	Mem	Mem	Mem	Mem (111 fetuses) Mus/Mem (1 fetus))

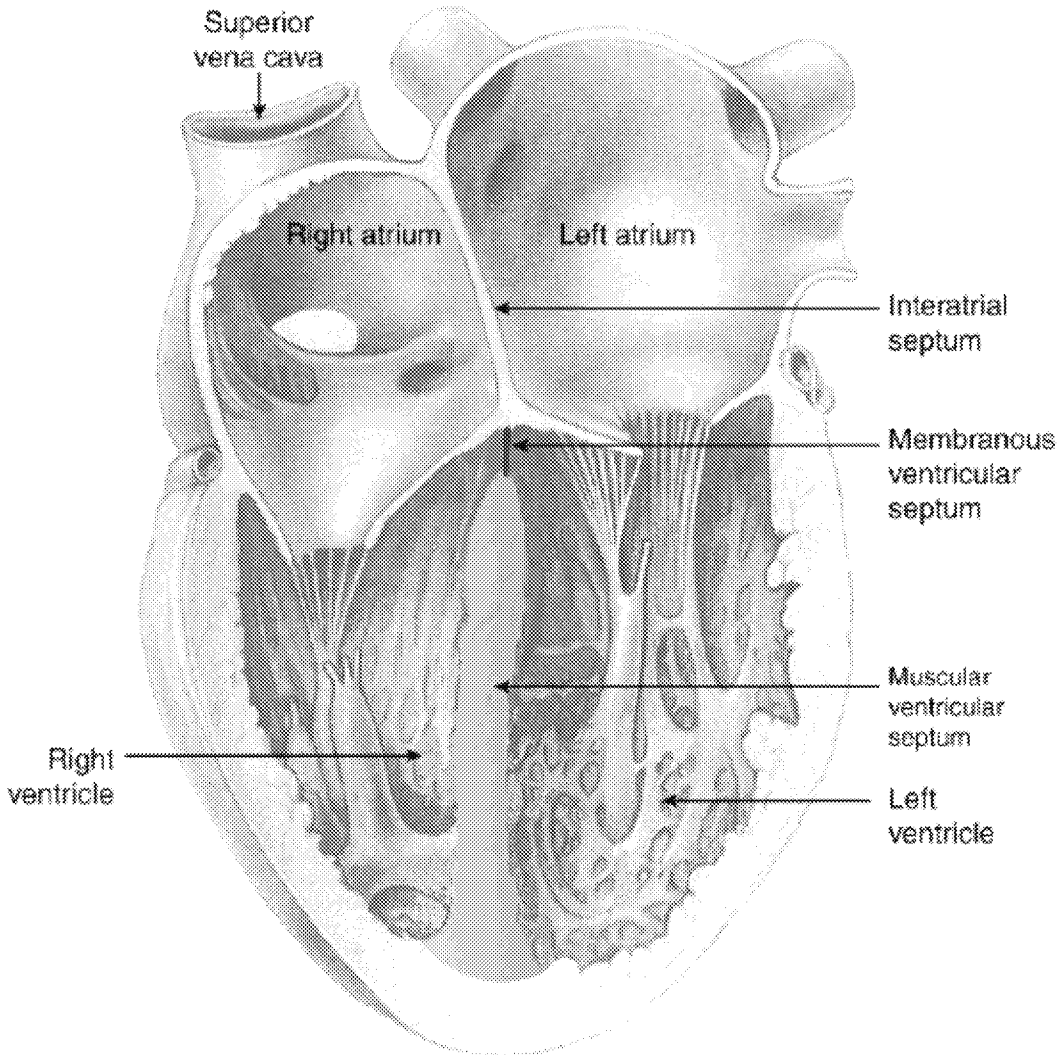
* One female had a very early total litter resorption and therefore had no available fetuses

** Significantly different from vehicle control group at p≤0.01

Mem = Membranous portion of ventricular septum

Mus/Mem = Muscular and membranous portions of ventricular septum

Location of Ventral Septal Defects



VSD Incidence in Control S-D Rats Using Enhanced Cardiac Evaluations

Study	Ventricular Septal Defects (%)
Haring, 1964[a]	3.2
Haring, 1966[a]	2.8
Inomata & Yasuda, 1971[b]	5.2
Inomata & Yasuda, 1971[b]	3.6
Solomon et al., 1997	2.4
Current Study[c]	2.4

[a] Hearts were embedded in paraffin, serially sectioned and examined by light microscopy

[b] Examined by a combination of Wilson freehand razor sections plus microdissection of the cardiac outflow tract

[c] Examined by fresh dissection

Health Risk Implications of Small (< 1 mm) VSDs in Rats and Humans

- VSDs in control and TCE-treated rats were all < 1 mm in size.
- A 2.4% incidence of < 1 mm VSDs present in untreated near-term SD rat fetuses was completely resolved at weaning (Solomon *et al.,* 1997).
- Trimethadione increased small VSDs in rat fetuses that were resolved at weaning while large VSDs were not, indicating closure of VSDs depends on the severity of the lesion at term (Fleeman *et al.*, 2004).
- Non-statistically significant formation of VSDs of < 1 mm in TCE drinking water treated rats is not an adverse health risk.

Comparison of TCE & TCA Drinking Water Plasma Levels to Other Inhalation or Oral Treatments

- TCE Non-Detects in drinking water dosing indicates parent TCE is not a dosimetrically plausible teratogen as postulated by Johnson *et al.* (2003).
- Higher TCE and TCA levels after inhalation and gavage doses indicates that an absence of cardiac malformations by these routes was not due to insufficient systemic TCE/TCA dosing.

Plasma TCE/TCA	Peak Plasma Concentration (µg/ml)				
	TCE				TCA
	Drinking Water[a] (0.25 & 1,000 ppm)	Drinking Water[b] (350 ppm)	Inhalation[b] (600 ppm, 4 hr/day)	Gavage[b] (2.3 mg/kg)	Gavage[c] (98 mg/kg)
TCE	ND	ND	24	0.26	----
TCA	1.1-1.2	2.8	13	25	201

[a] Current study, ND = Below LOD (50 ng/ml); TCA detected only 500 & 1000 ppm, not at 0.25 & 1.5 ppm.

[b] Fisher *et al.* (1989); TCE & TCA measured in pregnant GD21 SD rats.

[c] Larson & Bull (1992); TCA measured in adult male rats.

Conclusions: Risk Assessment Implications

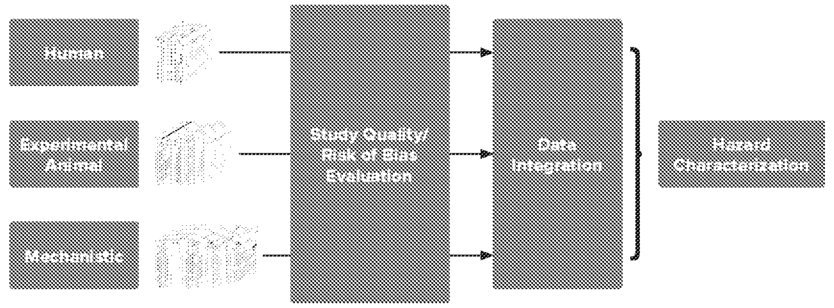
- TCE ingestion in drinking water at high concentrations of 1000 ppm (close to limit of water solubility) does not cause cardiac defects in rats.
- The findings of Johnson *et al.* (2003) are not reproduced in a high-quality GLP and regulatory guideline-consistent study.
- The drinking water plasma toxicokinetic data further enhance the conclusion that TCE is not a cardiac teratogen even under conditions of high-dose inhalation or gavage dosing resulting in substantially higher plasma TCE and TCA concentrations.

RELIABLE EVIDENCE INDICATES TCE IS NOT A CARDIAC TERATOGEN AFTER DRINKING WATER, INHALATION OR GAVAGE EXPOSURES

Comparison of NTP OHAT and USEPA TSCA Study Quality Criteria — Trichloroethylene (TCE) and Congenital Heart Defects (CHDs) as a Case Study

Urban J[1]; Wikoff D[2]; Suh M[3]; Britt J[4]; Harvey S[5]; Chappell G; Haws, L[1]

[1]ToxStrategies, Austin, TX 78759; [2]ToxStrategies, Asheville, NC 28804; [3]ToxStrategies, Mission Viejo, CA 92691; [4]ToxStrategies, Tallahassee, FL 32309; [5]ToxStrategies, Katy, TX 77494



Introduction

- Current TCE regulations are based on a single experimental animal study reporting an association between gestational exposure to trichloroethylene (TCE) and the development of congenital heart defects (CHDs) in offspring. This TCE-CHD association is controversial as it was not observed in the 11 other TCE developmental animal toxicology studies, including GLP studies specifically designed to repeat the single study that reported an association.
- Globally, systematic review is being implemented to facilitate hazard and risk evaluations. As part of developing best practices, various critical appraisal tools are being designed or refined to accommodate evidence bases which include many different types of studies (i.e., experimental animal, *in vitro*, and observational studies).
- Existing critical appraisal tools generally assess the internal validity of a study, as assessed by the risk of bias (RoB). With increasing use in toxicology, it is being recognized that other aspects of study quality are important on the individual study level (e.g., external validity or relevance). Some such aspects are considered in a newly released study quality tool issued by the USEPA-OPPT to facilitate the TSCA risk assessment systematic review process (USEPA, 2018). This tool is unique as it includes study quality metrics for mechanistic studies (e.g., *in vitro* studies).
- To date, only subsets of the TCE-CHD literature (human, animal) have been subjected to formal systematic critical appraisal (Wikoff et al., 2018).

Objective

To assess the impact of various systematic critical appraisal tools in the evaluation of the full evidence base (human, animal, and mechanistic data) for TCE-CHD.

Methods

Development of TCE-CHD Evidence Base (Literature Search):

- Using ad hoc searching and reference chasing, epidemiology, animal toxicology, and mechanistic studies were identified from recent comprehensive reviews conducted systematically (Makris et al., 2016; Wikoff et al., 2018). Additional PubMed and Embase searches were also conducted using the same search syntax utilized to capture relevant studies published since Wikoff et al. (2018). Searches were executed October 30, 2018.
- Mechanistic studies were categorized based on the assay type(s) to accommodate the TSCA study quality tool: *in vivo* (animals exposed), *in vitro* (cell culture, *in ovo*, *ex ovo*, *ex vivo*).

Critical Appraisal Tools (Table 1)

- **OHAT RoB:** Two study categories (animal and human) with defined, and reviewer refined, criteria for assessing bias (low, high; definite, probable). The RoB for the TCE-CHD human and animal literature is based on Wikoff et al. (2018).
- **TSCA Study Quality Evaluation:** Three study categories (human, *in vivo*, and *in vitro*) with specific evaluation and scoring metrics, each metric being scored on 1 of 4 criteria; overall study quality was determined by weighted scoring calculations and categorizations.
- **SciRAP:** Used in this effort to compare TSCA *in vitro* study quality results. Criteria evaluate the reporting and methodological quality, and relevance of *in vivo* and *in vitro* studies. Tool calculates a score for each category based on reviewer selection of several criteria.

Study Quality Assessment Procedure

- Pilot assessments: Independent review of subset of studies/experiments by two analysts for each of the three study categories. Decisions, interpretations, and refinements were documented.
- Quality assessments were conducted by two scientists with experience reviewing epidemiology (MS, JB), experimental animal (SF, JU), and mechanistic (GC, JU) studies. In cases of conflict, a third scientist (DW) was consulted to facilitate a consensus solution.

Data Integration and Body of Evidence Assessment

- Integration approach is based on OHAT (2015) and builds on that from Wikoff et al. (2018) to include mechanistic data and consider data quality output as determined by various appraisal tools.
- For mechanistic data, confidence-rating factors proposed by OHAT (2015) were considered: magnitude/potency, dose-response, consistency, directness, validity. Consideration was also given to the biological plausibility of data in the context of an adverse outcome pathway construct (which also relies on data from animal and human evidence streams to characterize adverse outcomes; in the case of TCE-CHD, adverse outcomes are limited as the majority of data suggest lack of such).

Table 1. Critical Appraisal Tool Study Quality Domains/Metrics

Study Type	NTP-OHAT RoB (2015)	USEPA OPPT TSCA (2018)	SciRAP (2018)
Epidemiology	**7 Questions covering 6 Bias Domains:** 1. Selection Bias (1) 2. Confounding Bias (1) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**15 Metrics covering 5 Study Elements:** 1. Study Participation (3) 2. Exposure Characerization (3) 3. Outcome Assessment (2) 4. Potential Counfounding/Variable Control (3) 5. Analysis (4)	[Not covered]
Animal Toxicology	**9 Questions covering 6 Bias Domains:**[*] 1. Selection Bias (2) 2. Performance Bias (2) 3. Attrition/Exclusion Bias (1) 4. Detection Bias (2) 5. Selective Reporting (1) 6. Other Bias (1 - covers statistics)	**24 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (3) 3. Exposure Characterization (6) 4. Test Organism/Test Model (3) 5. Outcome Assessment (5) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (2)	**53 Criteria covering 3 Study Elements:** 1. Reporting Quality (30) 2. Methodological Quality (18) 3. Relevance (5) [Not utilized in the current effort]
In vitro	[Not covered]	**25 Metrics covering 7 Study Elements:** 1. Test Substance (3) 2. Test Design (4) 3. Exposure Characterization (6) 4. Test Model (2) 5. Outcome Assessment (4) 6. Confounding/Variable Control (2) 7. Data Presentation and Analysis (4)	**42 Criteria covering 3 Study Elements:** 1. Reporting Quality (23) 2. Methodological Quality (15) 3. Relevance (4)

[*] OHAT RoB guidelines encourages reviewers to expand/modify/split questions to make relevant to PECO. For TCE-CHD case study, Wikoff et al. (2018) split 9 animal questions out to 14 questions.

Results

TCE-CHD Evidence Base

Table 2. TCE-CHD Literature

Study Type	# Relevant Published Studies	# Study Quality Assessments
Epidemiology	10	9
Animal Toxicology	11	12
Mechanistic	22	Total: 68 [Avg: 3.1 Assays/Study] In vivo: 5; In vitro (cell culture): 26; In vitro (in ovo): 21; In vitro (ex ovo): 3; In vitro (ex vivo): 7; In vitro (zebrafish): 5; Unknown model: 1
Total	**43**	**89**

Critical Appraisal of Epidemiological Data

- Overall study quality as assessed by the various tools was low for the epidemiological literature. Appraisal outcome was driven by limitations in study design and reporting particularly related to study participation, exposure assessment, and confounding.
- **Conclusion:** The nine studies comprising the human evidence base for TCE-CHD are of very limited study quality for risk assessment.

Table 3. Critical Appraisal of Human Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Epidemiology Studies			
Bove et al. (1995)/Bove (1996)	Cross-sectional (assumed exposure via public water)	Unacceptable (2x "4" scores)	Tier II
Brender et al. (2014)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Forand et al. (2012)	Ecological/Cross-sectional (assumed exposure via air)	High Quality (score=1.5)	Tier II
Gilboa et al. (2012)	Case-control (assumed exposure via air)	Unacceptable (1x "4" scores)	Tier II
Goldberg et al. (1990)	Pseudo-case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Lagakos et al. (1986)	Cross sectional (assumed exposure via public water)	Unacceptable (6x "4" scores)	Tier II
Ruckart et al. (2013)	Case-control (assumed exposure via public water)	Unacceptable (3x "4" scores)	Tier III
Tola et al. (1980)	Cohort (assumed exposure via air)	Unacceptable (9x "4" scores)	Tier II
Yauck et al. (2004)	Case-control (assumed exposure via air)	Unacceptable (4x "4" scores)	Tier II

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Experimental Animal Data

- Overall study quality as assessed by the various tools was medium to high for the experimental animal research. Appraisal outcome was largely driven by well reported and appropriate study design, consistent experimental conditions, and valid outcome methodologies.
- The Dawson et al. (1993)/Johnson et al. (2003) rat drinking water study was characterized as unreliable (poor study quality; high internal bias) by both OHAT and TSCA tools; common issues related to lack of concurrent controls, multiple vehicles within study groups, and unvalidated outcome assessment method.
- **Conclusion:** The majority of the animal evidence base for TCE-CHD [sans Dawson et al. (1993)/Johnson et al. (2003)] are amenable for risk assessment.

Table 4. Critical Appraisal of Animal Toxicology Studies Relevant to TCE-CHD Risk Assessment

Reference	Study Design	TSCA designation[1]	OHAT RoB Designation[2]
Oral Studies			
Cosby and Dukelow (1992)	Mouse - oral gavage GD 1-5, 6-10, or 11-15	Medium Quality (score=2.1)	Tier II
Dawson et al. (1993)/Johnson et al. (2003)	Rat - drinking water GD 1-22	Unacceptable (2x "4" scores)	Tier III
Fisher et al. (2001)	Rat - oral gavage GD 6-15	High Quality (score=1.5)	Tier II
Narotsky and Kavlock (1995)	Rat - oral gavage GD 6-19	Medium Quality (score=1.9)	Tier II
Narotsky et al. (1995)	Rat - oral gavage GD 6-15	Medium Quality (score=1.9)	Tier II
Inhalation Studies			
Carney et al. (2006)	Rat - whole body 6 hr/d, GD 6-20	High Quality (score=1.4)	Tier I
Dorfmueller et al. (1979)	Rat - whole body 6 hr/d, GD 1-20	Medium Quality (score=1.8)	Tier I
Hardin et al. (1981)a	Rat - whole body 7 hr/d, GD 1-19	High Quality (score=1.4)	Tier II
Hardin et al. (1981)b	Rabbit - whole body 7 hr/d, GD 1-22	High Quality (score=1.4)	Tier II
Healy et al. (1982)	Rat - whole body 4 hr/d, GD 8-21	Medium Quality (score=2.0)	Tier II
Schwetz et al. (1975)a	Rat - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I
Schwetz et al. (1975)b	Mouse - whole body 7 hr/d, GD 6-15	Medium Quality (score=1.8)	Tier I

[1] For OPPT scores, "high quality" studies <1.7, "medium quality" studies <2.3 and ≥1.7, "low quality" studies ≥2.3; any study with at least one metric score = 4 is automatically of "unacceptable quality".
[2] OHAT RoB Tiers as evaluated and reported in Wikoff et al. (2018)

Critical Appraisal of Mechanistic Datasets

- Pilot study of 10 experimental datasets using TSCA demonstrated that five study metrics commonly differentiated studies; these were defined as "Key Metrics." (**Table 5**)
- Quality rankings based on the TSCA tool varied by study model (**Figure 1**).
- Aspects that commonly differentiated studies within the TSCA tool included reporting on the preparation and storage of the test substance (Metric 8), elements of data analysis (Metrics 22 and/or 23), and reporting on cytotoxicity (Metric 24, only relevant to cell culture experiments) (**Figure 2**).
- Study quality categorizations were overall similar for the subset of experiments also assessed using SciRAP (**Table 6**).
- **Conclusion:** The majority of the mechanistic studies are not reliable for risk assessment. Traditional assessment parameters (e.g., magnitude, consistency) were not sufficient to facilitate conclusions for mechanistic data. Consideration of the type of outcome assessed (e.g., gene expression, *in ovo* development), the study model (e.g., chicken eggs, rat whole culture embryos, zebrafish larvae, human embryonic stem cells), as well as the plausibility of findings in a biological construct (e.g., adverse outcome pathway type of construct) were critical to integrating the evidence. The few mechanistic studies that were of sufficient quality were limited in their applicability due to heterogeneous models of questionable relevance to human physiology and exposure timing/dosing. Furthermore, the outcomes from these remaining studies were also inconsistent as it relates to outcome observations in mammalian species.

Table 5. Key Metrics Identified using TSCA Study Quality Metrics for TCE-CHD In Vitro Experiments

Metric No.	Metric Title	Metric Description
8	Preparation and Storage of Test Substance	Did the study characterize preparation of the test substance and storage conditions? Were the frequency of preparation and/or storage conditions appropriate to the test substance stability and solubility (if applicable)?
11	Exposure Duration	Was the exposure duration (e.g., minutes, hours, days) reported and appropriate for this study type and/or outcome(s) of interest?
16	Outcome Assessment Methodology	Did the outcome assessment methodology address or report the intended outcome(s) of interest? Was the outcome assessment methodology (including endpoints and timing of assessment) sensitive for the outcome(s) of interest (e.g., measured endpoints that are able to detect a true effect)?
22	Data Analysis	Were statistical methods, calculations methods, and/or data manipulation clearly described and appropriate for dataset(s)?
24	Cytotoxicity	Were cytotoxicity endpoints defined, if necessitated by study type, and were methods for measuring cytotoxicity described and commonly used for assessments?

Figure 1. TCE-CHD Mechanistic Studies by Model Type and Study Quality Category Based on TSCA Systematic Review Guidelines

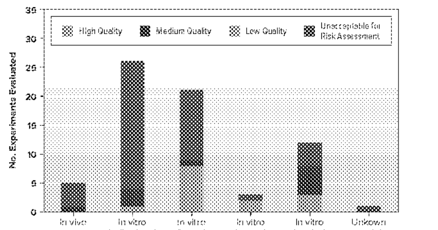


Figure 2. TSCA Study Quality Metrics Scored "Unacceptable" Across TCE-CHD Mechanistic Evidence Base

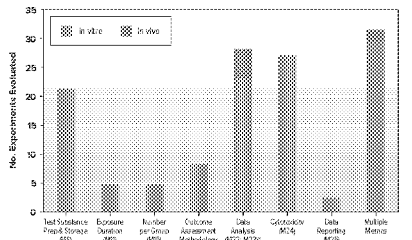
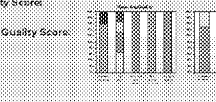
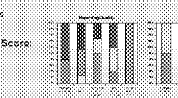
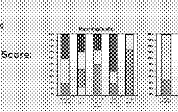


Table 6. Comparison of In Vitro Study Quality Evaluation Tools

Study	Experimental Model & Outcome	Study Quality Evaluation Outcome: TSCA	SciRAP	
Drake et al EHP (2006)	In ovo: Level of apoptosis in atrioventricular cushion and outflow tract sections of HH24 staged chick embryo hearts	**Score:** 1.3 **Interpretation:** High Quality Study **Consequence:** Retain in TCE-CHD evidence base for risk assessment	**Reporting Quality Score:** 78 out of 100 **Methodological Quality Score:** 85 out of 100 **Consequence:** None defined	
Harris et al (2018)	In vitro: Protein activity in transformed human liver cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 50 out of 100 **Methodological Quality Score:** 50 out of 100 **Consequence:** None defined	
Selmin et al (2005)	In vitro: Gene transcription levels in transformed rat heart cell line	**Score:** Not calculated, multiple metrics scored =4 **Interpretation:** Unacceptable for risk assessment **Consequence:** Remove from evidence base for TCE-CHD risk assessment	**Reporting Quality Score:** 43 out of 100 **Methodological Quality Score:** 46 out of 100 **Consequence:** None defined	

Body of Evidence Assessment

- Overall, there is higher confidence in the animal studies compared to human studies or mechanistic studies, based on the output of the various critical appraisal tools.
  1. Notably, the Dawson et al. (1993)/Johnson et al. (2003) study was determined to be unreliable by both appraisal tools. This emphasizes the likelihood that shortcomings in methodological and reporting aspects can explain the inconsistent findings of this study relative to the other 11 animal studies in the evidence base.
- Data Integration (**Figure 3**): Considered together, the available human, animal, and mechanistic study data support a lack of association between gestational TCE exposure and CHDs.
  1. Human studies → Low confidence in evidence stream associating in utero TCE exposure with increased risk of CHDs (similar to conclusions using OHAT RoB tool): Only a single study met TSCA quality criteria, and that was an ecological study.
  2. Animal studies → High confidence in evidence stream for TCE-CHD null hypothesis (i.e., no association of gestational TCE exposure and increased CHD risk): Only study to show dose response effect failed to meet TSCA study quality criteria.
  3. Mechanistic studies → Low confidence in evidence stream: inconsistency and relevance of outcomes and non-mammalian models are difficult to interpret given the lack of effect in experimental animal models (mammalian).

Figure 3. Data Integration: Evidence Stream Summaries and Integrated Conclusion

Conclusions and Confidence By Stream: Human Data	Animal Data	Mechanistic Data	Integrated Conclusion
Low to very low (inadequate) confidence to determine the potential for, or the direction of, an effect	+ Very high level of confidence supporting lack of association between TCE exposure (oral and inhalation) and CHD Single inconsistency explained by methological and reporting limitations*	+ Low confidence in the body of evidence to characterize CHD or mechanistic events potentially associated with CHD Complex interpretation given inconsistinecy in findings relative to other streams and limitations in relevance of non mammalian study models combined with limitations in study quality	→ Available evidence support a lack of association between gestational TCE exposure and CHD in humans

Conclusions

- Despite differences in the critical appraisal tools employed herein, consideration of study quality resulted in similar findings: the experimental animal studies offer the highest level of confidence. Both approaches deemed the Johnson et al. (2003) rat study unreliable for using in quantitative risk assessment.
- Given the consistent findings of experimental animal studies demonstrating a lack of TCE-CHD relationship, the utility of assessing and integrating the mechanistic data is limited, particularly considering the complexity of interpreting the relevance of diverse models (e.g., non-mammalian) and exposure paradigms (e.g., direct *in vitro* cell culture exposures extrapolate to high exposure concentrations in humans) utilized in a risk assessment context. Notably, in contrast to the rodent data, non-mammalian models (*in ovo*, zebrafish) provide the strongest evidence supporting TCE-CHD association. These models are heuristic tools useful for hypothesis development but are of highly questionable relevance for human health risk assessment.
- The use of multiple tools for evaluating the quality of study data across evidence bases can increase confidence in systematic review findings and provide an understanding of the practical application of available approaches.

Acknowledgements This work was funded by the American Chemistry Council.

References
Bove FJ. (1996). PMID: 8794538.
Bove FJ et al. (1995). PMID: 7717362
Brender JD et al. (2014). PMID: 25406847
Carney EW et al. (2006). PMID: 17066414.
Cosby NC & Dukelow WR. (1992). PMID: 1516784.
Dawson BV et al. (1993). PMID: 8473657.
Dorfmueller MA et al. (1979). PMID: 538767.
Drake VJ et al. (2006). PMID: 16759982.
Fisher JW et al. (2001). PMID: 11766123.
Forand SP et al. (2012). PMID: 22142966.
Gilboa SM et al. (2012). PMID: 22811060.
Goldberg SJ et al. (1990). PMID: 2358589.
Hardin BD et al. (1981). PMID: 7330632.
Harris AP et al. (2018). PMID: 29306027.
Healy TE et al. (1982). PMID: 7066129.
Johnson PD et al. (2003). PMID: 12611656.
Lagakos SW et al. (1986). DOI: 10.2307/2288982
Narotsky MG et al. (1995). PMID: 8529815.
Narotsky MG & Kavlock RJ. (1995). PMID: 7783251.
NTP-OHAT. (2015). https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html.
Ruckart PZ et al. (2013). PMID: 24304547.
Schwetz BA et al. (1975). PMID: 1135881.
Science in Risk Assessment and Policy (SciRAP). (2018). http://www.scirap.org/
Selmin O et al. (2005). PMID: 16142583.
Tola S et al. (1980). PMID: 7441393.
USEPA. (2018). EPA Document#740-P1-8001.
Yauck JS et al. (2004). PMID: 15390315.
Wikoff D et al. (2018). PMID: 29357719.